

IBM Applied Data Science Capstone Project by Coursera

Opening a new Coffee Café in Bengaluru, India

Girish A Koushik

July 2020

BUSINESS PROBLEM

- Coffee is one of the main beverages consumed highly (more than 5 times a day!) by people in South India.
- Bengaluru, being the Silicon Valley of India, is often visited and dwelled by people from various parts of the country, who come here seeking jobs.
- So, people are always searching for good coffee bars / coffee café / coffee shops across the city.
- Also, my personal inspiration comes from the fact that, my parents own a small coffee business and are looking to expand across the city.
- Business Question: In the city of Bengaluru, India, if a business owner is looking to open a new coffee café', where would you recommend them to open it?

DATA

- **Data Collection:**

1. List of neighborhoods in Bengaluru
2. Latitude and Longitude coordinates of the neighborhoods
3. Venue data related to coffee shops

- **Data Sources:**

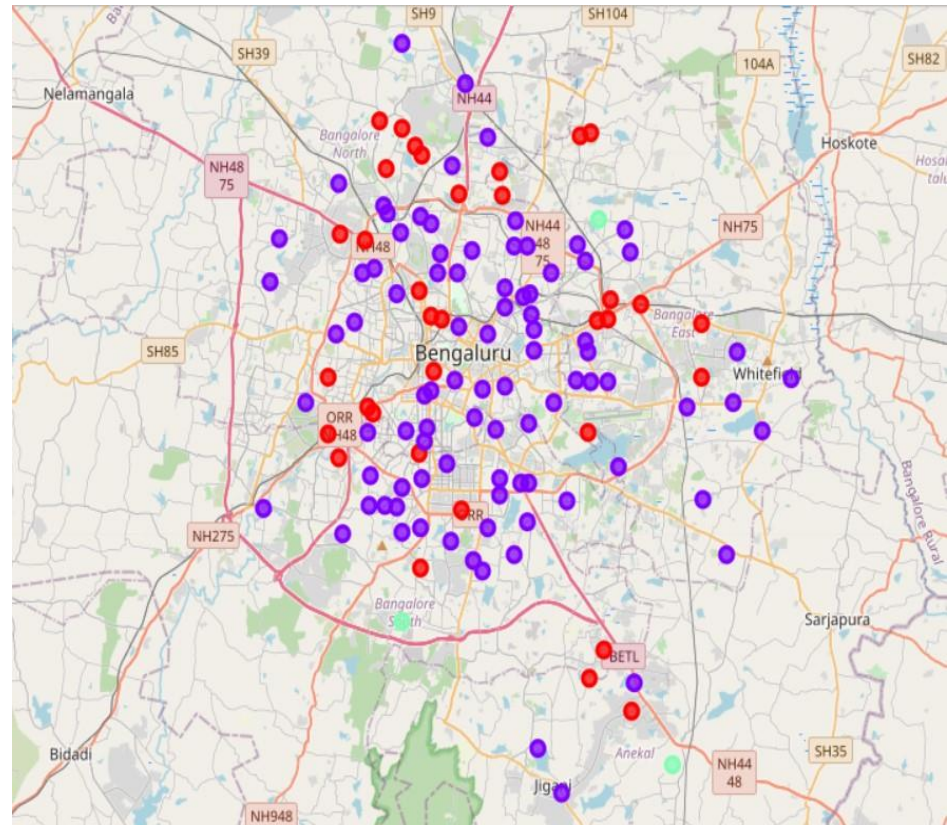
1. Wikipedia page for neighborhoods
(https://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Bangalore)
2. Geocoder package for latitude and longitude coordinates
3. Foursquare API for venue data

METHODOLOGY

- Scraping web data from Wikipedia page for neighborhoods list
- Obtain latitude and longitude coordinates using Geocoder
- Use Foursquare API to obtain venue data
- Group data by neighborhood, take mean of frequency of occurrence of each venue category in all the neighborhoods
- Filter venue category by coffee shops
- Perform clustering on the data by using k-means clustering
- Visualize the clusters in a map using Folium

3. Climate change and high seas

1. Cluster 0: Neighborhoods with moderate number of coffee shops.
2. Cluster 1: Neighborhoods with very low number to no coffee shops.
3. Cluster 2: Neighborhoods with high concentration of coffee shops.



OBSERVATIONS

- Most of the coffee shops are concentrated in the southern, eastern and central parts of the Bengaluru city.
- High concentration of coffee shops are found in cluster 2 whereas moderate number of coffee shops are found in cluster 0.
- Coffee shops in cluster 2 are likely suffering from intense competition due to large number of other coffee shops.
- Cluster 1 has very low number to no coffee shops in the neighborhoods and this presents a great opportunity to open new coffee shops there.

CONCLUSION

- This project implements the Data Science methodology i.e., Business Understanding → Data Requirements → Data Collection → Data Preparation → Modeling and Evaluation.
- This project provides recommendations to the relevant stakeholders i.e., coffee business owners and popular coffee chains regarding the best locations to open a new coffee café.
- Answer to the Business Question: The neighbourhoods in cluster 1 with low frequency of occurrence are the most preferred locations to open a new coffee café.

FUTURE RESEARCH

- Frequency of occurrence of coffee shops is the only feature used in this project.
- Other factors such as population, location of new corporate offices and data of different income groups can be highly useful.
- Also, small coffee shops already present may not be available in the Foursquare API.
- If it is possible to get the above data, then the accuracy of the prediction model can be highly improved.

THANK YOU!